A Bayesian Approach to High Resolution 3D Surface Reconstruction from Multiple Images

Robin D. Morris Peter Cheeseman Vadim N. Smelyanskiy
David A. Maluf
NASA Ames Research Center, MS 269-2, Moffett Field, CA 94035, USA
[rdm,cheesem,vadim,maluf]@ptolemy.arc.nasa.gov

Abstract

We present a radically different approach to the recovery of the three dimensional geometric and reflectance properties of a surface from image data. We pose the problem in a Bayesian framework, and proceed to infer the parameters of the model describing the surface. This allows great flexibility in the specification of the model, in terms of how both the geometrical properties and surface reflectance are specified. In the usual manner for Bayesian approaches it requires that we can simulate the data that would have been recorded for any state of the model in order to infer the model. The theoretical aspects are thus very general. We present results for one type of surface geometry (the triangular mesh) and for the Lambertian model of light scattering. Our framework also allows the easy incorporation of data from multiple sensing modalities.

1. Introduction

We present a radidcally different approach to the recovery of 3D geometry and surface reflectance information than those usually considered in computer vision, one that allows surface recovery from images taken under widely varying lighting conditions and using different cameras. This enables, for example, surface inference from images from different satellites, or inference using images from both satellites and planetary rovers. The two most commonly used conventional methods are shape-from-stereo and shape-from-shading.

In shape-from-stereo[7], correspondence matches are made between discrete points in the different images (using the epipolar constraint). Knowlege of the camera viewing geometry enables these matches to be used to recover a set of points in 3D space which lie on the surface. These points are then connected to form a representation of the surface.

In shape-from-shading[3] the gradients of the image intensities are directly related to the surface gradients. Integrating from a specified boundary condition enables a height value (more strictly, a distance from the camera value) to be associated with each pixel in the image.

Both of these methods have a number of drawbacks. In shape-from-stereo the density of the points is unknown a-priori and is dependent on the number of distinct point matches found. It can also be unclear how to join these points to form a surface.

In shape-from-shading the density of points is fixed at the image resolution. Shape-from-shading is difficult for surfaces where the surface reflectance properties are spatially varying. In both of these methods it is difficult to incorporate new observations into the surface estimation.

When the image formation process in linear, a Bayesian approach to surface reconstruction is given in [6].

In our approach we *begin* by specifying the surface model, both the geometrical aspects and the surface reflectance properties. Thus we may choose the level of detail of our model representation to suit our final purpose and on the basis of the data available. We may also refine the representation at a later time if more or higher resolution data becomes available. Commonly used models for the surface geometry include triangulated meshes and B-spline surfaces. In this paper we discuss only the triangulated mesh surface representation, and we limit the surface to be a height field, assigning the (x, y) coordinates of the vertices and learning only the z-values (heights).

Regarding how the surface reflects light, many standard models of light reflection are known, at increasing levels of complexity[5, 4]. The simplest model commonly used is that of Lambertian, or perfectly diffuse reflection. This model has a single parameter, the (wavelength dependent) albedo. More complex models specify further how the bi-directional reflectance function varies with the illumination and viewing geometry, and incorporate effects such as a specular component, the "hot spot" and models of surface roughness. Any parameterised model can be used in our framework; here we present results only for the Lambertian

model.

Thus we have posed the surface reconstruction problem as the problem of estimating the parameters of a surface model from image data. The estimation of model parameters from data is best solved using Bayesian methods[1], and the approach taken is given in the following sections.

2. Bayesian estimation for surface model parameters

The specific form of the general surface model that we consider here is parameterised by a set of heights and a set of albedos, and the triangulation of these heights to form the surface. We do not consider here the estimation of the camera parameters, nor the parameters of the illumination incident on the surface; these are assumed known. The estimation of the camera parameters will be considered in a forthcoming paper.

Bayes theorem states that

$$p(h,\rho|I_1 \dots I_n) \propto p(I_1 \dots I_n|h,\rho)p(h,\rho)$$

where h,ρ is the vector of heights and albedos and I_i is the image data. This states that that the posterior distribution of the heights and the albedos is proportional to the likelihood – the probability of observing the data given the heights and albedos – multiplied by the prior distribution on the heights and albedos.

Consider first the likelihood. We make the usual assumption that the differences between the observed data and the data synthesised from the model have a zero mean, Gaussian distribution, and also assume that the images comprising the data are conditionally independent. This gives

$$p(I_1 \dots I_n|h,\rho) \propto \prod_i \exp\left(-\frac{\sum(I_i - \hat{I}_i(h,\rho))^2}{2\sigma_e^2}\right)$$

where $\hat{I}_i(h,\rho)$ denotes the image synthesised from the model, σ_e^2 is the noise variance and the summation is over the pixels.

The prior distribution is also assumed to be Gaussian, and determined by the integral value of the squares of the surface curvature, $\int c(x,y)dxdy$, where

$$c(x,y) = \left(\frac{\partial^2 h}{\partial x^2}\right)^2 + \left(\frac{\partial^2 h}{\partial y^2}\right)^2 + 2\left(\frac{\partial^2 h}{\partial x \partial y}\right)^2$$

The partial derivatives are approximated by finite differences of the height and albedo values. The coefficients of $h(i,j) \times h(i+p, j+q)$ from $c(x,y)$ summed over the surface give the entries in the prior inverse covariance matrix.

$$p(h,\rho) \propto \exp\left(-[h\ \rho]\Sigma^{-1}[h\ \rho]^T/2\right)$$

This prior is placed over the height variables, h, but the albedos are only defined over the range $[0-1]$. Because of this we put the gaussian prior for the albedos over transformed variables, where

$$\rho \rightarrow \log(\rho'/(1-\rho')) \tag{1}$$

Forming the prior covariance matrix in this way ensures that it is positive-definite.

Consider the negative log-posterior. For a single image we have

$$L(h,\rho) \propto \frac{\sum(I - \hat{I}(h,\rho))^2}{\sigma_e^2} + \begin{bmatrix} h-h_0 \\ \rho-\rho_0 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} h-h_0 \\ \rho-\rho_0 \end{bmatrix} \tag{2}$$

and this is a nonlinear function of h,ρ. The MAP estimate is that value of h,ρ which minimises $L(h,\rho)$.

In the case of images with no shadows or visible occlusions which we consider here, the log-posterior is in general unimodal and gradient methods can be applied for minimising $L(h,\rho)$. We linearise $\hat{I}(h,\rho)$ about the current estimate, h_0,ρ_0 and replace $\hat{I}(h,\rho)$ by $\hat{I}(h_0,\rho_0) + \mathbf{D}\begin{bmatrix} h-h_0 \\ \rho-\rho_0 \end{bmatrix}$ where $\mathbf{D}$ is the matrix of derivatives evaluated at h_0,ρ_0.

$$\mathbf{D}_{ij} = \frac{\partial \text{pixel } i}{\partial \text{height (or albedo) } j}$$

The minimisation of equation 2 then becomes the minimisation of the quadratic form

$$L'(h,\rho) = \begin{bmatrix} h-h_0 \\ \rho-\rho_0 \end{bmatrix}^T \left(\Sigma^{-1} + \frac{\mathbf{D}\mathbf{D}^T}{\sigma_e^2}\right) \begin{bmatrix} h-h_0 \\ \rho-\rho_0 \end{bmatrix} - \frac{(I - \hat{I}(h_o,\rho_0))}{\sigma_e^2}\mathbf{D}\begin{bmatrix} h-h_0 \\ \rho-\rho_0 \end{bmatrix} + \frac{\sum(I-\hat{I}(h_0,\rho_0))^2}{\sigma_e^2}$$

We mimisie this quadratic form using the conjugate gradient method. This finds the minimum of the local linear approximation. At the minimum we recompute $\hat{I}$ and $\mathbf{D}$ and minimise $L(h,\rho)$ (equation 2) iteratively. We found that typically convergence occurs in four to five iterations.

Thus to find the MAP estimate requires that we can compute $\hat{I}$ and $\mathbf{D}$ for any values of h,ρ. We discuss this computation in some detail in the next section. Here it is sufficient to note that while forming $\hat{I}$ using only object space computation (see section 3) is computationally expensive, we can compute $\mathbf{D}$ at the same time for little additional computation. This makes the process described above a practical one. Convergence can also be accelerated by using a multigrid approach.

At convergence we compute the new inverse covariance matrix, $(\Sigma^{-1})' = \Sigma^{-1} + \mathbf{D}^T\mathbf{D}/\sigma_e$. This is then used as the *prior* inverse covariance matrix when new image data of the same surface is obtained, enabling a recursive update and integration of data recorded at different times. The posterior inverse covariance matrix gives information about the uncertainty of the estimated surface.

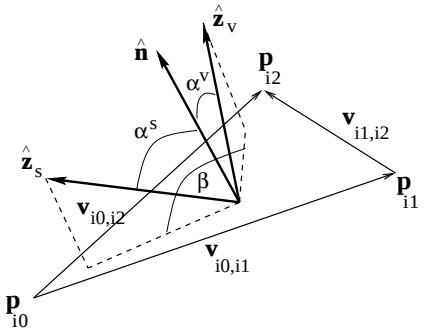

Figure 1. Geometry of the triangular facet, illumination direction and viewing direction

3. Forming the image and the derivative matrix

The task of forming an image, $\hat{I}$, given a surface description, h, ρ, and camera and illumination parameters is the area of computer graphics known as rendering[2]. However, because of their target application (producing images for *people* to look at), current computer graphics systems which render from triangulated meshes make one fundamental assumption which leaves them unsuitable here. This assumption is that, when projected into the image plane, each triangle making up the surface is much larger than a pixel, so that the approximation that any given pixel is illuminated by light from only one triangular facet is reasonable. This allows images which are visually appealing to be computed quickly, but can lead to aliasing artefacts at the edges of the triangles, and produce inaccurate images if the projected triangles are much smaller than the pixels. In the computer graphics literature these are known as *image space* algorithms.

For our system we implemented a renderer for triangular meshes which does all computation in *object space*. When the light from a triangle is projected into a pixel, its contribution to the brightness of that pixel is weighted by the fraction of the area of the triangle which projects into the pixel. This produces perfectly anti-aliased images and allows an image of any resolution to be produced from a mesh of arbitrary density, as required when the system performing the surface inference may have no control over the image data gathering.

Figure 1 shows a single triangular facet from the surface model. $\hat{\mathbf{z}}_s$ is the unit vector in the direction to the illumination, $\hat{\mathbf{z}}_v$ the direction to the camera (viewing direction) and $\hat{\mathbf{n}}$ is the surface normal. Let $\mathcal{I}_s$ be the intensity of the direct illumination and $\mathcal{I}_a$ be the intensity of the ambient light. The flux reflected from this facet into the spatial angle $\Delta\Omega$ in the viewing direction is

$$\Phi = E(\alpha_s) r(\alpha_s, \alpha_v, \beta) \Delta\Omega \tag{3}$$

where $E(\alpha_s)$ is the incident flux and $r()$ is the bi-directional reflectance function. $\Delta\Omega = d/R^2$ where R is the distance from the facet to the sensor and d is the area of the aperture.

$$E(\alpha_s) = \rho A(\mathcal{I}_s \mathbf{n} \cdot \hat{\mathbf{z}}_s + \mathcal{I}_a), \quad A = \frac{1}{2}\mathbf{v}_{i0,i1} \times \mathbf{v}_{i1,i2} \tag{4}$$

The derivative of Φ with respect to ρ is clear from equations 3 and 4 and the logarithmic transformation given in equation 1. The derivative with respect to the z-component of $\mathbf{p}_{i0}$ is more complex, being made up of a number of components.

$$\frac{\partial \Phi}{\partial z_{i0}} = \frac{d}{R^2}\left(\frac{\partial E}{\partial z_{i0}} r + E \frac{\partial r}{\partial z_{i0}}\right)$$

where

$$\frac{\partial r}{\partial z_{i0}} = \frac{\partial r}{\partial \alpha_s}\frac{\partial \alpha_s}{\partial z_{i0}} + \frac{\partial r}{\partial \alpha_v}\frac{\partial \alpha_v}{\partial z_{i0}} + \frac{\partial r}{\partial \beta}\frac{\partial \beta}{\partial z_{i0}}$$

If shadows are present on the surface and visible occlusions are present in the image then these must be taken into account when computing the derivatives. In fact, these non-local derivatives are very informative as to the shape of the surface.

For the case of Lambertian reflectance considered here, $r() = \cos\alpha_v$, and with no shadows or occlusions the derivative with respect to the height is

$$\frac{\partial \Phi}{\partial \mathbf{z}_{i0}} = \frac{1}{2}\rho(\mathbf{p}_{i2} - \mathbf{p}_{i1}) \times \hat{\mathbf{z}} \cdot \mathbf{g}$$

where

$$\mathbf{g} = \mathcal{I}_s(\hat{\mathbf{z}}_v \cos\alpha_s + \hat{\mathbf{z}}_s \cos\alpha_v - \hat{\mathbf{n}} \cos\alpha_s \cos\alpha_v) + \mathcal{I}_a \hat{\mathbf{z}}_v$$

If the triangle does not project entirely within a pixel then the area of overlap must be considered. When forming the image, as mentioned above, the flux Φ must be weighted by the fraction of the triangle's area which projects into the pixel. When computing the contribution to the derivative of a pixel from the vertices of that triangle, the derivative of the area fraction with respect to the z_i must be included.

Note that the derivatives in this section assume a perfectly focused image. If blurring is present then both the synthesised image formation and the derivative computations are modified.

4. Results

Figure 2 shows a low-resolution image of a synthetic surface (the image is 45×45 pixels). Four such images, with differing lighting and camera orientations were produced and used as the data images I. It is easy to show that at least three images of each point on the surface are required.

Informally this can be seen by noting that we must recover two components of the surface normal and the albedo at each point.

Starting from a mesh with all zero heights and all albedos set to 0.5, the conjugate gradient scheme described above was used to infer the surface shown in figure 3. The surface is of dimension 64×64 heights and the same number of albedos. The curvature based prior was also used. Note that because of the near vertical viewpoint there is only a very weak dependence on the mean distance to the surface. To overcome this we assumed known boundary conditions. Also note that this is a dense triangluation – when projected into the pixel grid of figure 2 many triangles fall into one pixel. Thus we infer a super-resolved surface – a pixel lying on the rim of the crater does not imply a planar region in the inferred surface, rather, we infer a surface where highly curved regions may project into a single pixel.

Error maps at the end of 5 minimisations are shown in figure 4. Note the vertical scales compared with figure 3. The reconstruction is extremely accurate, with most errors being in the regions of high curvature. The images synthesised from the infered surface are visually indistinguishable from the data images (see figure 2).

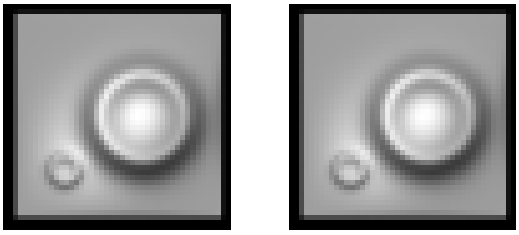

Figure 2. Low resolution image of the real surface (left) and the inferred surface (right)

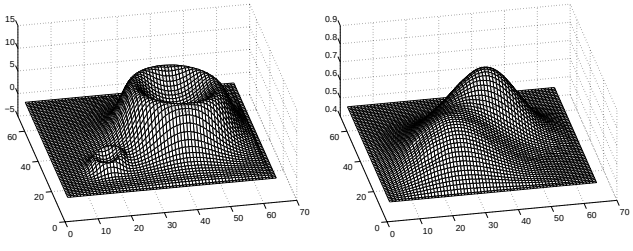

Figure 3. Heights (left) and albedos (right) for the inferred surface

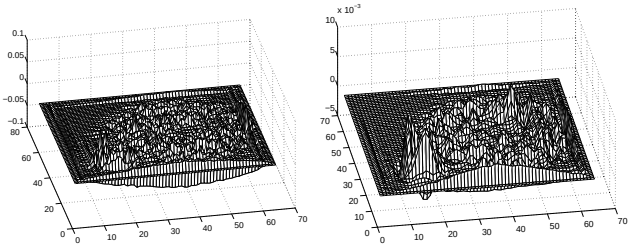

Figure 4. The errors between the inferred and the true surface (Heights (left) and albedos (right)

5. Conclusions and future extensions

We have developed a very general framework for the inference of general surface geometry and reflectance models from image data, where the model choice is determined by the physical properties of the surface we wish to infer. We have demonstrated that for the case of a triangulated surface and Lambertian reflectance the parameters of a surface model, namely the heights and albedos, can be inferred from a set of image data. We have developed a framework that allows easy inclusion of future data observed from the same surface, and easy incorporation of data from other sensing modalities.

Future developments will include the addition of the ability to correctly compute the image and its derivatives when shadows and visible occlusions are present. Extensions to other surface representations and other surface topologies and more realistic reflection functions will also be studied. Limits to the accuracy of the surface reconstruction will also be explored.

References

[1] J. Bernardo and A. Smith. *Bayesian Theory*. Wiley, Chichester, New York, 1994.

[2] J. Foley, A. van Dam, S. Finer, and J. Hughes. *Computer Graphics, principles and practice*. Addison-Wesley, 2nd ed. edition, 1990.

[3] B. Horn and M. Brooks. *Shape from Shading*. MIT Press, 1989.

[4] S. Nayar and M. Oren. Visual appearance of matte surfaces. *Science*, 267:1153–1156, February 1995.

[5] W. Rees. *Physical principles of remote sensing*. Cambridge University Press, 1990.

[6] D. Wolf. A bayesian reflection on surfaces. In *Maximum Entropy and Bayesian Methods*. Kluwer, 1998.

[7] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report No 2273, INRIA, Sophia Antipolis, 1994.